Project proposal
Speech classifier using SVM linear model/ Naive Bayes/ Neural Network
By Group 3

Goal

Evaluate if public opinion about a product is positive or negative when it is expressed in a way that does not include numerical ratings such as stars or points - for example, a set of tweets about a new game. To achieve this goal we plan to use and compare several models trained on a set of product reviews that do include numerical ratings (retrieved from Amazon), which will evaluate the tweets' sentiment on a scale from 1 to 5. We will be using an SVM linear model, a Naive Bayes classifier, and a basic neural network; we will compare results from all three.

Training

We will use a portion (~80%) of an Amazon customer reviews dataset to train the models. The dataset contains a few million reviews and their respective star ratings and can be found at this link: https://www.kaggle.com/bittlingmayer/amazonreviews.

Testing

The remainder of the Amazon dataset will be used to test the models, such that we can confirm it works and has a high enough accuracy. We will use this information to guide us when we perform the evaluation.

Evaluation

For the evaluation, we will choose a dataset of tweets related to a certain product (or several products depending on the amount of data available). We will then use our trained and tested model to assign positivity/negativity ratings to the tweets, which do not contain a numerical expression of the poster's sentiment like Amazon reviews do. In this way, we hope to be able to assign a numerical rating to these product opinions found on social media. We will do this with each of the three options (SVM, NB, neural network) and compare the results.